

1st Project

Traffic Crash in Tempe, AZ Analysis

Data Collection: Already available

Visualization

- Spatial Visualization
- Heatmaps (density of accidents in Tempe)
- Plot number of accidents over years/months/days
- Visualize how injury/fatality counts correlate with driver age, gender, alcohol/drug usage, or violation types
- Compare holidays and non-holidays number of accidents

Optional: make a dashboard

Model

- Classification Models: Predict injury severity or fatality occurrence based on conditions like weather, time, collision type, and driver behavior
- Regression Models: quantify effects of variables like alcohol use, light conditions, etc., in relationship with total fatality, total injury,
- Spatial Scan to identify areas at higher accident risk

Ethics

- Maintain privacy and confidentiality standards
- Research purpose only, avoid unnecessarily stigmatizing locations

2nd Project

Stock Returns and Revenue Dynamics Analysis of Leading Tech and E-Commerce Firms

5 big companies, and 5 small companies (same sector)

Big companies: Amazon, Apple, Google, Walmart, Tesla

Small companies: Etsy, Zynga, MongoDB, GameStop Corporation, Wayfair

How to collect data

- Use yahoo finance library to get the daily stock returns of each company from 2014 to 2024
- Then, web-scraped the annual revenue of those companies (SEC EDGAR website)

Visualization

- Bar charts or line charts of daily/cumulative returns
- Bar charts or line charts of annual revenue
- Do stock prices and revenue trends side-by-side
- Correlation heatmaps

Optional: make a dashboard

Modeling:

- **Volatility model**
- **Regression model, observing the relationship between stock returns and revenue**
- **ARIMA for predicting future prices**

Ethics: Use only public, non-personal data, for research purpose only